33rd Annual Department of Defense Cost Analysis Symposium

Specifying Probability Distributions From Partial Information on their Ranges of Values

Paul R. Garvey
Chief Scientist
The Economic and Decision Analysis Center
The MITRE Corporation

Main Points

- In systems engineering, probability distributions whose values are uncertain must often be specified by expert technical opinion
- In the absence of meaningful (or historical) data, expert opinion is often the only way to quantify a variable's uncertainty
- Instead of assigning a single subjective probability to an event, subject experts often find it easier to describe a function that depicts a subjective distribution of probabilities
- Such a distribution is sometimes called a subjective probability distribution
- Because of their nature, subjective probability distributions can be thought of as "belief functions" – mathematical representations of a subject expert's best professional judgment in the distribution of probabilities for a particular event

MITRE

Main Points

- When formulating a subjective probability distribution to describe the uncertainty in a variable's value, subject experts often prefer specifying a range that contains most, but not all, possible values
- Thus, there is a nonzero probability that values will occur outside the expert's specified range
- One strategy for specifying a subjective probability distribution involves the direct assessment of the distribution's fractiles
- Another strategy involves assigning a subjective probability to a subinterval of the range of the distribution function
- This talk illustrates these strategies in the context of new formulas for specifying beta, uniform, and triangular distributions from only partial information about their ranges of values; four cases will be represented

Subjective Probabilities and Distribution Functions

- In systems engineering, probabilities are often used to quantify uncertainties associated with a system's design parameters (e.g., weight), as well as uncertainties in cost and schedule
- For reasons previously mentioned, quantifying these uncertainties is often done in terms of subjective probabilities
- Subjective probabilities are those assigned to events on the basis of personal judgment; they measure a person's degree of belief that an event will occur
- Subjective probabilities are most often associated with one-time, non-repeatable, events – those whose probabilities cannot be objectively determined from a population of outcomes developed by repeated trials, observations, or experimentation

Subjective Probabilities and Distribution Functions

- Subjective probabilities cannot be arbitrary; they *must adhere* to the three fundamental axioms of probability

  Axiom 1 states the probability of any event is a nonnegative number in the interval zero to unity. Axiom 2 states that $P(W)$ is equal to unity, where the sample space W is sometimes referred to as the *sure* or *certain event*. Axiom 3 states for any sequence of mutually exclusive events, the probability of at least one of these events occurring is the sum of the probabilities associated with each event A_i. In axiom 3, this sequence may also be finite.

- For instance, if an electronics engineer assigns a probability of 0.7 to the event *"the number of gates for the new processor chip **will not exceed** 12,000,"* it must follow that the chip **will exceed** 12,000 gates with probability 0.3
- Subjective probabilities are conditional on the state of the person's knowledge which changes with time
- To be credible, subjective probabilities should only be assigned to events by subject experts - persons with significant experience with events similar to the one under consideration

MITRE

Subjective Probabilities and Distribution Functions

- In addition, the rationale supporting the assigned probability must be well documented
- Imposing a disciplined approach to defining and documenting subjective probabilities lessens the chance of encountering the "village watchman"

Sir Josiah Stamp* once said...

"The government are very keen on amassing statistics. They collect them, raise them to the n-th power, take the cube root, and prepare wonderful diagrams. But one must never forget that every one of these figures comes in the first instance from the village watchman, who puts down what he damn pleases."

* President of the Bank of England during the 1920s.

Specifying Probability Distributions From
Partial Information on their Ranges of Values…Beta Distribution Case

- The beta distribution has long been the distribution of "choice" for specifying subjective probability distributions
- It can take a wide variety of forms, as seen in the figures below

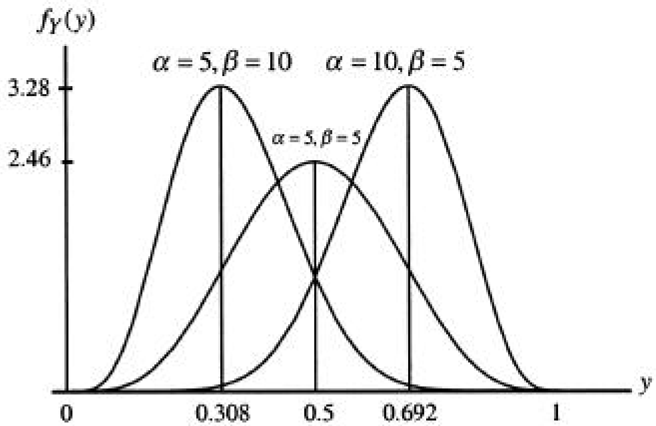


Figure 4-9. A Family of Standard Beta Probability Density Functions

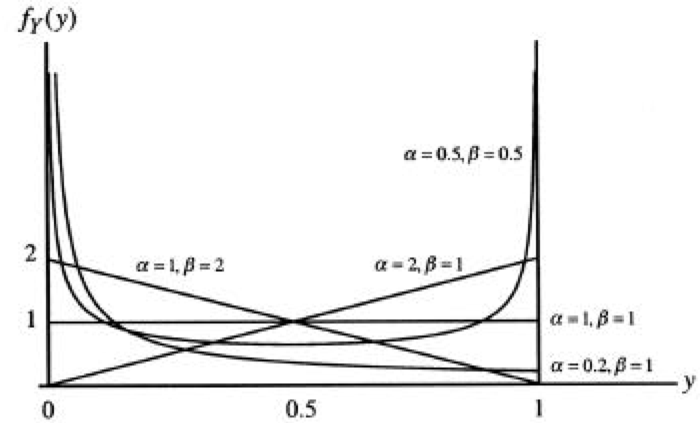


Figure 4-10. More Standard Beta Probability Density Functions

MITRE

Specifying Probability Distributions From

Partial Information on their Ranges of Values…Beta Distribution Case

- The beta distribution can be used to describe a random variable (e.g., weight, lines of code, schedule) whose range of possible values is bounded by an interval of the real line
- A random variable X is said to have a *nonstandard beta distribution* if its probability density function is given by

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\dfrac{x-a}{b-a}\right)^{a-1} \left(\dfrac{b-x}{b-a}\right)^{b-1} & a < x < b \\ 0 & \text{otherwise} \end{cases} \qquad (4\text{-}8)$$

where a and b ($a > 0$ and $b > 0$) determine the shape of the density function and $\Gamma(a)$ is the gamma function of the argument a

- A random variable X with density function given by Eq 4-8 will be implied by the expression $X \sim Beta(a, b, a, b)$

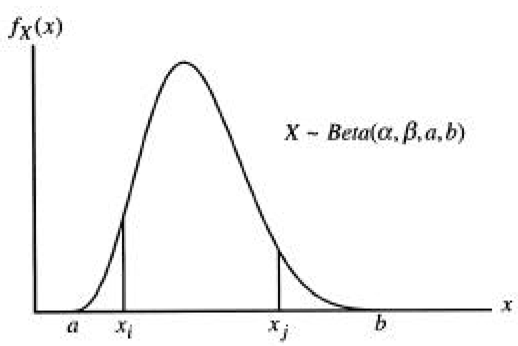
Specifying Probability Distributions From

Partial Information on their Ranges of Values…Beta Distribution Case

- A random variable *Y* is said to have a *standard beta distribution* if its probability density function is given by

$$f_Y(y) = \begin{cases} \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(y)^{a-1}(1-y)^{b-1} & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (4\text{-}9)$$

- A random variable *Y* with density function given by Eq 4-9 will be implied by the expression $Y \sim Beta(a, b)$
- The transformation of $X \sim Beta(a, b, a, b) \longrightarrow Y \sim Beta(a, b)$

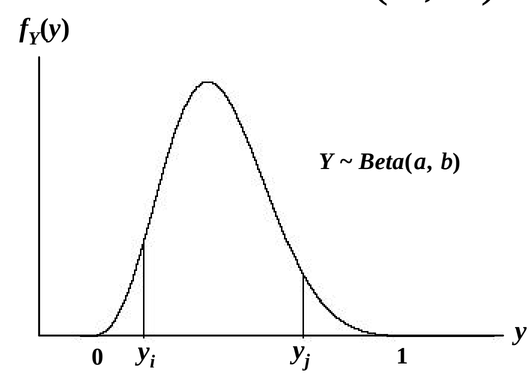


$$y = (x-a)/(b-a)$$

$\longrightarrow$

$Prob(Y \le y_i) = i = Prob(X \le x_i)$

$Prob(Y \le y_j) = j = Prob(X \le x_j)$



Specifying Probability Distributions From

Partial Information on their Ranges of Values…Beta Distribution

- Case 1: Specify a nonstandard beta distribution for the random variable X given the shape parameters a and b and any two fractiles x_i and x_j , where $0 \leq i < j \leq 1$; an illustration of this case is presented in figure 4-20
- Purposes: To determine the minimum possible value of X, denoted by a, and the maximum possible value of X, denoted by b, where $X \sim Beta(a, b, a, b)$. To compute $E(X)$ and $Var(X)$ from the specified distribution

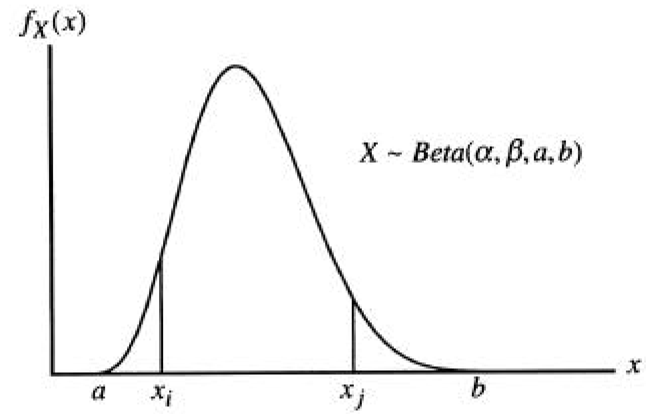


Figure 4-20. An Illustrative Beta Distribution — Case 1

The value x_a is called the a-fractile of X if

$$P(X \leq x_a) = a$$

Specifying Probability Distributions From
Partial Information on their Ranges of Values…Beta Distribution

- We are given a and b and *any* two fractiles x_i and x_j; since

$$Prob(Y \le y_i) = i = Prob(X \le x_i)$$
$$Prob(Y \le y_j) = j = Prob(X \le x_j)$$

where x_i and x_j are from $X \sim Beta(a, b, a, b)$; y_i and y_j are from $Y \sim Beta(a, b)$ and

$$y_i = (x_i - a)/(b - a)$$
$$y_j = (x_j - a)/(b - a)$$

Minimum value of X $\longrightarrow$ $$a = \frac{x_i y_j - x_j y_i}{y_j - y_i} \qquad (4\text{-}48)$$

Maximum value of X $\longrightarrow$ $$b = \frac{x_j(1 - y_i) - x_i(1 - y_j)}{y_j - y_i} \qquad (4\text{-}49)$$

MITRE

Specifying Probability Distributions From
Partial Information on their Ranges of Values…Beta Distribution

- Practical Example

  Suppose *I* represents the uncertainty in the number of delivered source instructions (DSI) for a new software application. Suppose a team of software engineers judged 100,000 DSI as a reasonable assessment of the 50th percentile of *I* and a size of 150,000 DSI as a reasonable assessment of the 95th percentile. Furthermore, suppose the distribution function in figure 4-21 was considered a good characterization of the uncertainty in the number of DSI.

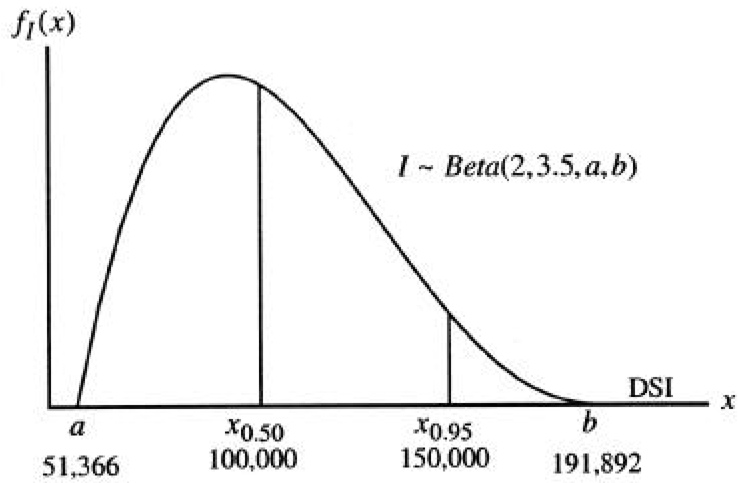


Figure 4-21. Beta Distribution for Example 4-11

Specifying Probability Distributions From
Partial Information on their Ranges of Values…Beta Distribution

- Practical Example (concluded):

We are given that *Prob*(I ⊠ $x_{0.50}$ = 100,000) = 0.50 and *Prob*(I ⊠ $x_{0.95}$ = 150,000) = 0.95.

Since a = 2 and b = 3.5, the *standard beta distribution* is $Y \sim$ Beta (2, 3.5); from this we can determine the fractiles $y_{0.50}$ and $y_{0.95}$. Using a tool such as *Mathematica*® we have $y_{0.50}$ = 0.346086 and $y_{0.95}$ = 0.70189. Substituting these values into equations 4-48 and 4-49 provides the minimum and maximum possible

$$a = \frac{(100000)0.70189 - (150000)0.346086}{0.70189 - 0.346086} = 51366$$

$$b = \frac{150000(1 - 0.346086) - 100000(1 - 0.70189)}{0.70189 - 0.346086} = 191892$$

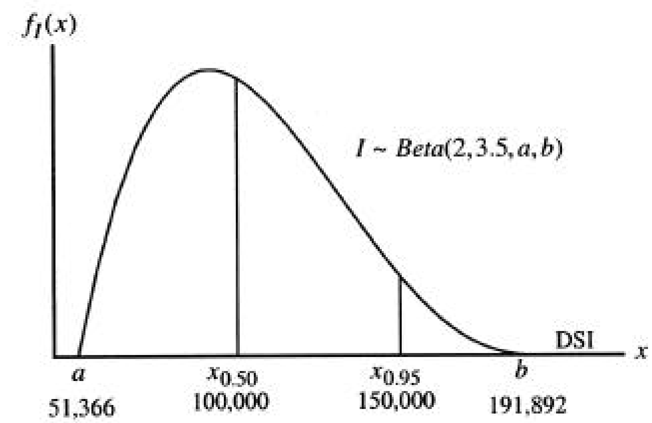


Figure 4-21. Beta Distribution for Example 4-11

Use the *Mathematica*® routine Quantile[BetaDistribution[2,3.5],*k*] where *k* is equal to 0.50 and 0.95, respectively, to obtain the two "*y*" fractiles.

Specifying Probability Distributions From

Partial Information on their Ranges of Values…Uniform Distribution

- Case 2: Specify a uniform distribution for the random variable X given the subinterval $a \leq x \leq b^{\circ}$ and α where a is the minimum possible value of X and

  $$b^{\circ} < b, \text{ and } \alpha = Prob(a \leq X \leq b^{\circ})$$

  an illustration of this case is presented in figure 4-22
- Purposes: To determine the maximum possible value of X, denoted by b. To compute $E(X)$ and $Var(X)$ from the specified distribution
- Required Information:
  Assessments of α and the endpoints of the subinterval $a \leq x \leq b^{\circ}$

ans. The maximum value of X is

$$b = a + (b^{\circ} - a)/\alpha$$

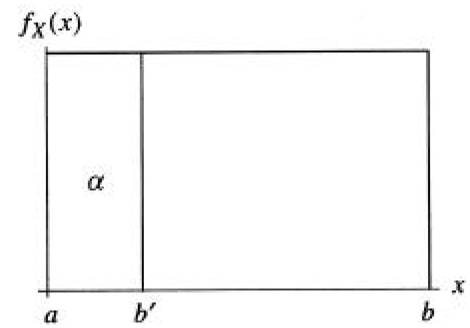


Figure 4-22. An Illustrative Uniform Distribution — Case 2

MITRE

Specifying Probability Distributions From
Partial Information on their Ranges of Values…Uniform Distribution

- Case 3: Specify a uniform distribution for the random variable X given the subinterval $a^{\circ} \le x \le b^{\circ}$ and α where

$$a < a^{\circ}\,,\ b^{\circ} < b, \text{ and } \alpha = Prob(a^{\circ} \le X \le b^{\circ})$$

  an illustration of this case is presented in figure 4-24
- Purposes: To determine the minimum possible value of X, denoted by a, and the maximum possible value of X, denoted by b. To compute $E(X)$ and $Var(X)$ from the specified distribution
- Required Information:
  Assessments of α and the endpoints of the subinterval $a^{\circ} \le x \le b^{\circ}$; furthermore, assume $a^{\circ} - a = b - b^{\circ}$

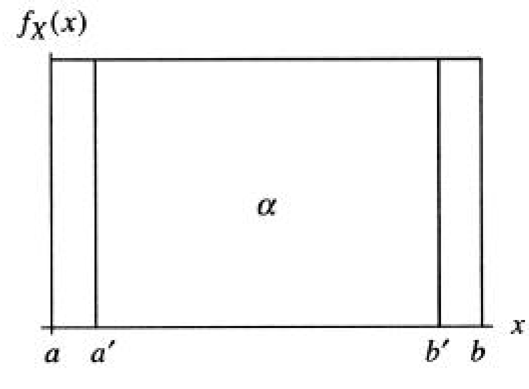


Figure 4-24. An Illustrative Uniform Distribution — Case 3

From Garvey, P. R., "*Probability Methods for Cost Uncertainty Analysis-A Systems Engineering Perspective*", published by Marcel Dekker, Inc., 270 Madison Avenue, New York, NY, 10016-0602, 2000; http://www.dekker.com/e/p.pl/8966-0.

MITRE

Specifying Probability Distributions From
Partial Information on their Ranges of Values…Uniform Distribution

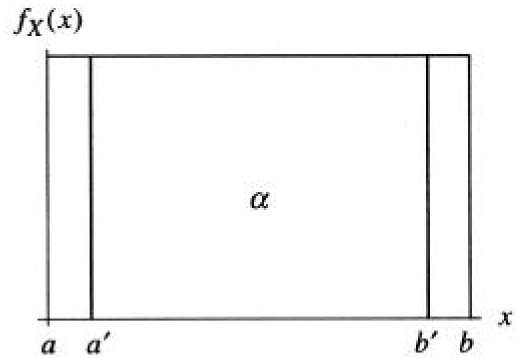


Figure 4-24. An Illustrative Uniform Distribution — Case 3

If $P(a' \le X \le b') = \alpha < 1$, the minimum and maximum possible values of X are

$$a = a' - \frac{1-\alpha}{2\alpha}(b' - a') \qquad (4\text{-}53)$$

$$b = b' + \frac{1-\alpha}{2\alpha}(b' - a') \qquad (4\text{-}54)$$

Notice that $a' - a = b - b'$. Furthermore, for this case we have

$$P(a \le X < a') = P(b' < X \le b) = \tfrac{1}{2}(1-\alpha)$$

Specifying Probability Distributions From

Partial Information on their Ranges of Values…Triangular Distribution*

- Case 4: Specify a triangular distribution for the random variable X given m, the subinterval $a^{\circ} \leq x \leq b^{\circ}$, and α where

$$a < a^{\circ}\ ,\ a^{\circ} \leq m \leq b^{\circ}\ ,\ b^{\circ} < b, \text{ and } \alpha = Prob(a^{\circ} \leq X \leq b^{\circ}\ )$$

  an illustration of this case is presented in figure 4-26

- Purposes: To determine the minimum possible value of X, denoted by a, and the maximum possible value of X, denoted by b. To compute $E(X)$ and $Var(X)$ from the specified distribution

- Required Information: Assessments of α and the endpoints of the subinterval $a^{\circ} \leq x \leq b^{\circ}$, where $a^{\circ} \leq m \leq b^{\circ}$

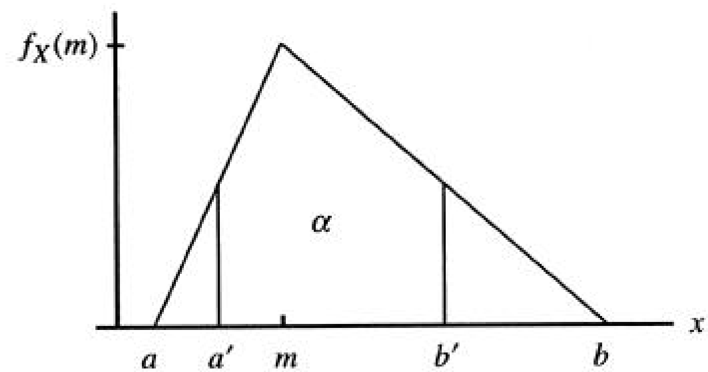


Figure 4-26. An Illustrative Triangular Distribution — Case 4

* This case was developed by Dr. C. C. Cho, The MITRE Corporation, Bedford, MA

Specifying Probability Distributions From
Partial Information on their Ranges of Values…Triangular Distribution

Assuming that

$$\frac{P(X \le a')}{P(X \ge b')} = \frac{P(X \le m)}{P(X \ge m)}$$

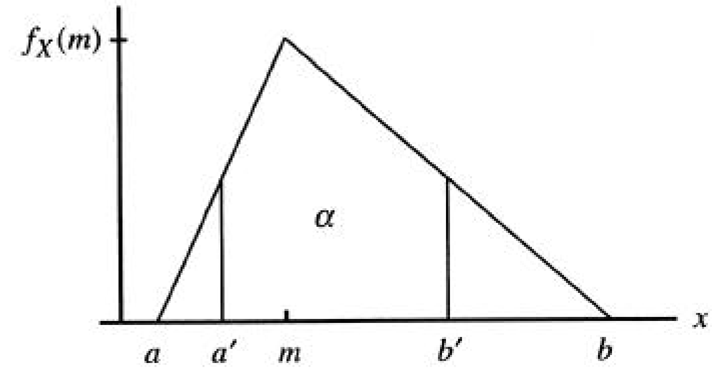


Figure 4-26. An Illustrative Triangular Distribution — Case 4

If $P(a' \le X \le b') = \alpha < 1$, the minimum and maximum possible values of X are

$$a = m - \frac{m - a'}{1 - \sqrt{1 - \alpha}} \quad (4\text{-}55)$$

$$b = m + \frac{b' - m}{1 - \sqrt{1 - \alpha}} \quad (4\text{-}56)$$

Summary

“That’s all Folks!”...

http://members.home.net/asb/Loony.htm
http://landru.i-link-2.net/debbie/looney/porky.html